SYMMETRY, MODEL REDUCTION, AND QUANTUM MECHANICS.
Abbreviated: Models and quantum mechanics.

Inge S. Helland
Department of Mathematics,
University of Oslo,
P.O. Box 1053 Blindern,
N-0316 Oslo, Norway
E-mail: ingeh@math.uio.no*

Abstract

Taking several statistical examples, in particular one involving a choice of experiment, as points of departure, and making symmetry assumptions, the link towards quantum theory developed in Helland (2005a,b) is surveyed and clarified. The quantum Hilbert space is constructed from the parameters of the various experiments using group representation theory. It is shown under natural assumptions that a subset of the set of unit vectors of this space, the generalized coherent state vectors, can be put in correspondence with questions of the kind: What is the value of the (complete) parameter? - together with a crisp answer to that question. Links are made to statistical models in general, to model reduction of overparametrized models and to the design of experiments. It turns out to be essential that the range of the statistical parameter is an invariant set under the relevant symmetry group.

1 Introduction.

Statistical modelling is at the core of much applied science. Nevertheless, the very concept of a model is sometimes debated, and is in fact debatable.

*AMS 2000 subject classification. Primary 62A01. Secondary 81P10, 62B15. Key words and phrases: Hilbert space, model reduction, quantum mechanics, question and answer, state, symmetry.

Drum and McCullagh (1993) attack 'the megalomaniacal strategy of fitting a grand unified model, supposedly capable of answering any conceivable question that might be posed'. Likewise, applied oriented books and papers like Burnham and Anderson (1998) and Bozdagan (1987) are very sceptical to the existence of a 'true model' in moderately complicated situations.

Some, like Breiman (2001), take the extreme position of rejecting more or less totally the concept of a parametric model and seek other ways to do data analysis. In this paper we will keep the model concept as a central one, and we will look upon a model as a tool, and not least as a way to give a language through which we can describe nature. Everybody agrees that a model, if used, should be chosen carefully from subject matter knowledge. We will claim that in certain situations it may be equally important to choose a model by taking into account what can be done statistically. A very rich model may give a good conceptual background, but it may also make inference impossible. In such cases it is not always the most fruitful attitude to look upon a narrower model as an approximation; it may be just the particular reduced model which is adequate for inference.

The perhaps surprising conclusion of this paper is that such a view of modelling is not only relevant for applied statistics; it is also a view that seems to be consistent with quantum mechanics, that is, if we follow the approach to the foundation of the theory which has recently been proposed by Helland (2005a,b). Below we will also give a summary of this approach from a statistician's point of view.

Quantum mechanics is a science that has reached its success through a very abstract kind of modelling. Our claim is that it is possible to make a link to this formal world from the modelling process and the inference process which is usual in statistics.

In fact, the time seems to be ripe for taking such a wide perspective. In recent years, Bayesian ideas have entered strongly into quantum physics (Fuchs, 2002, Bovens and Hartmann, 2003). This development goes together with a change in interpretation from a basically ontological to an epistemological point of view, that is, from emphasizing how the world 'is' to emphasizing how the world can be interpreted by us. From a statistical point of view, the physicist's version of Bayesianism - although deep philosophical ideas are involved - may be seen as somewhat unusual, however, since the distinction between parameter and observation has not been much stressed in this part of the literature. A historical reason for this latter attitude may be that most measurement apparata used in physics are very accurate, so there has seldom been a need to put forward a statistical measurement model.

2 An example.

Examples with dice, urns and card games abound in elementary probability, but not many such examples include the aspect of choice of focus under limited resources. The following example is inspired by a simpler example in Taraldsen (1995), and can act as a model for how we see some of the mechanisms behind quantum mechanics. It is too simpleminded, though, to illustrate all these mechanisms.

EXAMPLE 1. The following description is to be considered as a potential model behind data which are introduced later. Let a robot R be able to handle an apparatus A consisting of 1) an ordinary die and 2) a pack of six cards, among these one and only one ace. We instruct the robot to choose between 1) and 2) from some arbitrary unknown mechanism. If 1) is chosen, the die is thrown two times, and the information on which of the throws gives a one (result 1) or not one (result 0) is stored by the robot. If 2) is chosen, two cards are drawn randomly, and the information on which of the draws is an ace (result 1) or not (result 0) is stored by the robot.

Now assume that we are forced to read the result of this experiment through a one-bit computer. We can instruct the robot what to feed into this computer.

In experiment a the result of the first of the two throws/draws is reported, while in experiment b the robot reports 1 if there is at least one result equal to 1 in the two throws/draws, otherwise 0 is reported.

This can be repeated. It is assumed that the choice of die versus card pack (which is unknown to us) remains constant during these repetitions. But if we try to give the robot a new set of instructions during the series of experiments, the whole apparatus A is destroyed and replaced by a new one of the same kind. In particular, the robot makes a new choice between 1) and 2). So only one of the experiments a or b is permitted with the same apparatus.

Let us place ourselves in the rôle of observers/ experimenters in this experiment. We then only observe a series of bits, and this is all information we have in addition to knowing which instructions we have given to the robot. The rest, the die and the card deck must only be looked upon as models capable to describe our observations.

In experiment a we are able to estimate the probability of result 1, which may be taken as 1/6 in the model above, but which may have the possibility to be different in a refinement of the model. If we only perform

this experiment, we can satisfy ourselves with a very simple model: There is some bit-generating device such that the probability of 1 is 1/6.

In experiment b the frequency of 1's will approach 11/36 if a die was thrown, 16/36 if a card pack was used, so the experimenter is able to distinguish between these two possibilities. Again, if only experiment b was performed, some relatively simple model may be put forward to explain the result.

Now assume that we repeat the whole experiment several times, alternating between experiment a and experiment b. Then we get some patterns which to begin with may be difficult to understand, but after a while, we may put up with a total model involving a die and a pack of cards as above. We may study the series of experiments b closer, and find that the robot has a constant probability of choosing a die, or the frequency may alternate, which requires a more sophisticated model.

Returning to a single experiment when the overall model has been established, two different kinds of inference concerning A may be made according to whether experiment a or b has been chosen, but one experiment excludes the other. The underlying model for the experimental setting is comparatively complicated, but the probability model for each experiment, which is the basis for inference, is a simple binary one. This is a way of thinking that we claim may be useful in applied science: Although we have a complicated overall model, a simpler model may be useful for data analysis. The point here and elsewhere is that we have limited data.

One should be careful in drawing too many analogies to such a rather simple-minded example, but at least some resemblance to certain quantum mechanical situations may be seen. Readers knowing quantum theory may think of the double slit experiment, of measuring position or momentum for a particle, or of measuring spin components in different directions. In all these cases there is a choice of experiment, and the different experiments exclude each other. A natural implication of the present paper, is that in all these quantum mechanical cases it may be useful to have an underlying model. In the present paper, a concrete such model will only be proposed for the case of particle spin, but see the discussion.

The underlying model may in general be seen as related to some hidden parameter. There has been long debate on the rôle of hidden variable models in quantum mechanics, but since the paper of Bell (1966) there seems to be an agreement that such models can not be excluded altogether by a classical argument with a related intention given by von Neumann (1932). In this paper we will concentrate on what we will call hidden total parameters, mathematical variables which never can take a physical value. Such hidden

parameters, through being connected to a fairly flexible model concept, can be given a different status than the ontological hidden variables which, say, have been proposed as a basis for quantum physics by Bohm (1952).

The setting of Example 1 can be refined in many different directions. From a statistical point of view the most general refinement will consist of one thought experiment with data y and total parameter ϕ and some underlying limitation through which we are given portions of the data y^a or y^b according to our choice. The models for inference will then contain a parameter $\lambda^a = \lambda^a(\phi)$ for y^a and the parameter $\lambda^b = \lambda^b(\phi)$ for y^b.

3 Reduced models and inference.

In analogy to Example 1, let us have some mental model of a physical or other kind of system $\mathcal{S}$. Some part of this model will be taken as known; the rest constitutes what we will call a total parameter ϕ.

Assume that it is possible to make manipulations of $\mathcal{S}$, and that we after this can make some observations y^a. We let the letter a generically denote the choice of manipulations together with a relevant choice of observation window. We will assume that y^a is maximal, given a, in the sense that all data that can be obtained under this data window, are included.

Now by combining the mental model with what is known about the observational process (measurement apparatus), we find a statistical model $\mathrm{p^a}(\cdot|\lambda^{\mathrm{a}})$ for y^a. The parameter λ^a will be assumed to be a function of ϕ, and since y^a is maximal, also λ^a will be maximal in the sense that it is not a proper function of another experimental parameter. In principle λ^a could also be a function of other parameters connected to the measurement process, but it is important for us that we can disregard such complications. Let Λ^a be the space over which the parameter λ^a varies.

We will at the outset concentrate on Bayesian inference in this paper; but the symmetry assumptions of the next section will ensure that there is a large degree of agreement in conclusions between the Bayesian and frequentist point of view. The prior on λ^a will be chosen from symmetry, and this gives a posterior probability distribution $\pi^a(\cdot|y^a)$. We use standard inference from this.

4 Further examples; symmetry.

The model reduction from ϕ to λ^a can be of many kinds; some are related to sample space restrictions, some to unobservable quantities.

Example 2. Let the 'true' regression model of y upon a scalar variable x be a polynomial of high degree with regression coefficients $\phi = (\beta_0, \beta_1, ..., \beta_p)$. Assume now that the possible values of x consist of some equally spaced numbers $x_0, x_1, ..., x_q$ with $q < p$. Then the model can always without loss of generality be reduced to a polynomial of degree q with parameters $\lambda = (\beta_0, \beta_1, ..., \beta_q)$.

Example 3. Let our model consist of some solid equilateral triangle with corners A, B and C embedded inside a sphere which is non-transparent except for three windows 1, 2 and 3, equally spaced along the equator of the sphere. The centre of the triangle coincides with the centre of the sphere, and the three corners are on the sphere.

Let ϕ denote the position of the triangle, and let λ^a be the corner of the triangle which is closest to the window a. It is assumed that inference can be made on λ^a - and on λ^a only - through some observation y^a made from window a $(a = 1, 2, 3)$.

This example has some features in common with Example 1: There is an underlying model, which in the simplest case here is just the triangle, but it could also have been a more complicated figure. The experimenter has to make a choice: Which corner to measure. We assume that the different choices exclude each other. Given the choice a, there is a measurement model, a probability distribution of y^a, given λ^a. The additional feature of this example is that there is a symmetry between the different experiments.

Example 4. (Helland, 2005a). Four drugs A, B, C and D are being compared with respect to the expected recovery time μ they induce on patients with a certain disease. Let $\phi = (\mu_A, \mu_B, \mu_C, \mu_D)$. There are relatively few patients available, so one concentrates on getting information on the sign of the difference between each μ and the mean of the others, for instance $\lambda^A = \text{sign}[\mu_A - (\mu_B + \mu_C + \mu_D)/3]$. We will assume that there an incomplete block type design where accurate information can be obtained about one or a few such λ's.

One of our points is that the situation in these and other examples, where the underlying model is only indirectly related to observations, may make a natural opportunity to focus also upon other mathematical elements for a foundation for statistical inference than the usual measure theory and asymptotic theory. An obvious candidate here is given by the elements of

group theory and symmetry considerations that are relevant for inference. A review of this area is given in Helland (2004).

In fact, in most of the above examples and in many other examples we can make symmetry assumptions. In the rest of this paper we will assume that there is a transformation group G acting upon the total parameter space $\Phi = \{\phi\}$. Technically we will assume that the total parameter space can be given a topology such that it is locally compact, and that weak conditions (Helland, 2004, 2005b) hold which ensure that it can be given a measure ν which is right invariant under the group G (; i.e., $\nu(Eg) = \nu(E)$ for all $g \in G$ and all Borel sets $E \subset \Phi$. As in the above papers we place group actions to the right of the element it acts upon.)

The symmetry group may have several essential properties. First, it is important that the total parameter space is closed under the actions of G. This may be obviously true in most cases, but not always so when G is induced by another group acting upon the sample space.

Example 2 (continued). Why don't we usually propose models of the kind $\mathrm{E}(y) = \beta_0 + \beta_1 x + \beta_3 x^3$ in polynomial regression? One answer is that we want our class of models to be invariant under the actions $x \to x+a$. These translations induce group actions on the (total) parameter space which are easy to find, but cumbersome to write down. It is important that the parameter space is closed under this group.

An *orbit* of the group is any subset of Φ which is closed under the actions of the group. If there is only one orbit, so that any point in the space can be reached from any other point, the group is called *transitive.* This will be the ordinary situation in our setting, which we will see later.

A further important property is whether or not the group actions in G, as acting upon Φ in a natural way induce group actions on the image space of the parameters $\lambda^a = \lambda^a(\phi)$. For given a, this will always be the case if $\lambda^a(\phi_1) = \lambda^a(\phi_2)$ implies $\lambda^a(\phi_1 g) = \lambda^a(\phi_2 g)$ for all $g \in G$. If this property holds, we say that $\lambda^a(\cdot)$ is a *permissible* parameter. If it does not hold, we can always make it to hold by going from G to a subgroup G^a (Helland, 2005b). In fact, this way of constructing subgroups corresponding to each of several complementary parameters, seems to be fundamental for our understanding of quantum mechanics, as explained later.

Example 3 (continued). Let G be the group of rotations of the triangle, or what is equivalent, since we concentrate on the corners, the group of permutations of the three corners A, B and C. Then the parameters

$\lambda^a(\cdot)$ are not permissible, but they can be made permissible by going to a subgroup.

Here is a proof for λ^1: Let the positions ϕ be reduced to the possible permutations of the corners: ABC, CAB, BCA, ACB, CBA and BAC. It is enough to produce a counterexample: Let g be the permutation which exchanges the first and the last letter. Then $\lambda^1(ABC) = \lambda^1(ACB) = A$, while $\lambda^1(ABCg) = C$ and $\lambda^1(ACBg) = B$. The subgroup G^1 will here consist of the translations $ABC \to ABC$, $ABC \to BCA$, $ABC \to CAB$.

Finally, it is important that the group G^a really is relevant for statistical inference under the reduced model, which requires that the range Λ of λ^a is invariant under the actions of this group. We will make the stronger assumption that the range constitutes a single orbit of the group, that is, that G^a is transitive on Λ^a.

Assumption A. *The range Λ^a of λ^a is an orbit under the group G^a.*

For our development of quantum mechanics, we will also make an assumption which is not satisfied in Example 3, but which holds in the example of the following Section:

Assumption B. *The groups G^a generate G.*

5 Electron spin.

The spin component of an electron can be measured in any direction a, and it will always take one of the values -1 or +1. In our approach the spin itself can be modelled through a total parameter ϕ, a spin vector in three dimensional space whose direction gives the spin axis, and where the norm of ϕ gives the speed of rotation. Recall that we consider such total parameters merely as mathematical variables, not capable of taking any physical value.

The experimenter chooses a direction a, takes $\lambda^a(\phi) = \text{sign}(a \cdot \phi)$, and performes an experiment with λ^a as parameter. In analogy with the examples above, we can imagine some physical mechanism connected to the electron implying that only information about λ^a can be obtained once the direction a has been chosen.

The natural group G is the group of rotations of the vector ϕ, possibly taken together with a change in the norm of the vector. In any case, the parameter $\lambda^a(\cdot)$ is not a permissible function of the total parameter, since two vectors having the same value of the parameter will not have the same

value after a rotation. The maximal group G^a with respect to which λ^a is permissible, consists of all rotations around the axis a plus a 180^{o} rotation around any axis perpendicular to a. To this we may add norm changes, but since these don't affect λ^a, we concentrate on rotations.

Further discussion of this modelling approach to electron spin is given in Helland (2005b).

6 States as questions plus answers.

In the situations above we have a choice of experiment a, but once this choice is made, we have an ordinary inference situation with parameter λ^a and data y^a, and we assume an ordinary parametric model.

In the situation where a group G is defined on the total parameter space, it follows from Assumption A that the reduced group G^a is transitive on the range of λ^a. Then a unique right invariant measure under G^a can be defined, and we take this as an objective prior for Bayesian inference. The Bayes estimator is equivalent to the best equivariant estimator under quadratic loss, and the Bayes credibility intervals coincide numerically with the frequentist confidence intervals (Helland, 2004).

The conclusion of the experiment can be given as a posterior measure $\pi^a(\cdot|y^a)$ on the parameter λ^a. In the setting above this is equivalent to the confidence distribution of Schweder and Hjort (2003); see Helland (2004).

This conclusion can be formulated as a question plus an uncertain answer: The question lies in the choice of focus a, more precisely: What is the value of λ^a? The answer is given as a measure.

In the ideal case we have a perfect experiment, and the answer will be crisp: $\lambda^a = \lambda^a_0$. Such ideal experiments will in practice only be possible when the parameter space is discrete.

In any case we will define a state as the conclusion made about a system from such an experiment, that is, the choice of experiment a together with the crisp or uncertain result.

7 Quantum mechanics and group representations.

Elementary quantum mechanics can according to our view be seen as a tool for making decisions about a discrete parameter. The ordinary formal foundation is different, however: A pure state in quantum theory is defined as a unit vector in some fixed separable Hilbert space connected to the system in question. An observable is defined as a given operator on this Hilbert space.

This means that the eigenvectors of these operators are possible state vectors, and the corresponding eigenvalues are the values of the state variable given by this operator. For parts of what follows, but not for everything, we will expect that the reader has some familiarity with basic quantum mechanics in a way that can be found in any textbook. But in fact, logically it should be possible to follow the discussion below by just taking at face value the very brief description of the theory just given.

We will approach the formal world of quantum theory from our statistical point of departure. As a part of this development it turns out to be very useful to introduce the concept of a *group representation*.

A group representation is defined mathematically as a homomorphism from a given group to the space of operators upon some fixed vector space. This means that each group element g corresponds to an operator $U(g)$ on some Hilbert space, and the different operators satisfy $U(g)U(h) = U(gh)$. It follows that the identity group element is mapped into the identity operator, and that $U(g^{-1}) = U(g)^{-1}$. We will assume that the space upon which the operators act, is a complex vector space, and that the representing operators are unitary, so that $U(g)^{-1} = U(g)^{\dagger}$. This can always be assumed. In the finite-dimensional case, where $U(g)$ is a square matrix, the action $\dagger$ denotes transposition together with complex conjugation. In general $\dagger$ is defined through $(U^{\dagger}v, w) = (v, Uw)$ in terms of the scalar product of the Hilbert space. A unitary operator U then satisfies $(Uv, Uw) = (v, w)$ for all v, w.

If the underlying vector space is rich enough, one can say that the operator $U(g)$ characterizes the group elements g much in the same way as a characteristic function characterizes a probability measure.

In this paper we will concentrate first on a more concrete special case. Our fixed vector space will always be the space $\mathrm{L}^2(\Phi, \nu)$ of square integrable functions of ϕ under the right invariant measure ν associated with the basic group G, or a subspace of this space. We will then first confine ourselves to the *regular representation* where $U(g)$ is defined by

$$U(g)f(\phi) = f(\phi g). \tag{1}$$

It is easy to see that this defines a homomorphism and therefore a group representation: $U(gh) = U(g)U(h)$. It is also easy to see that the operator given by (1) is unitary.

There is a large general theory on group representations; see for instance Diaconis (1988) and James and Liebeck (1993).

8 The group representation approach to the quantum mechanical Hilbert space.

We will now return to statistics and to complementary choices of experiments, and use this setting to approach quantum mechanics. Of course, not all experimental situations described above may be expected to lead to the quantummechanical formalism. It turns out that this formalism only results when there is a symmetry between the various choices of experiment. In analogy with the treatment in Helland (2005b) this may be formulated as follows:

ASSUMPTION C. *For each pair of experiments a, b there is an element g_{ab} of the basic group G which induces a correspondence between the respective parameters: $\lambda^b(\phi) = \lambda^a(\phi g_{ab})$. The group elements $\{g_{ab}\}$ form a subgroup of G with $g_{ab}g_{bc} = g_{ac}$.*

As in Helland (2005b) this may be formulated informally as $\lambda^b = \lambda^a g_{ab}$. We will assume that the elements g_{ab} together with one fixed subgroup G^a generate the full group G, which will follow if the subgroups G^a together generate G.

Assumption C has consequences for the regular group representations. First, a few definitions:

DEFINITION 1. *a) $\mathbf{H}^a$ is the subset of $\mathrm{L}^2(\Phi, \nu)$ which consists of functions of the form $f(\phi) = \tilde{f}(\lambda^a(\phi))$, where $\tilde{f} \in \mathrm{L}^2(\Lambda^a, \nu^a)$, with ν^a being the invariant measure on Λ^a under the group G^a.*

b) $U^a(\cdot)$ is the regular representation of the group G^a on $\mathrm{L}^2(\Phi, \nu)$.

c) $U(\cdot)$ is the regular representation of the group G on $\mathrm{L}^2(\Phi, \nu)$.

Then we have:

THEOREM 1. *a) $\mathbf{H}^a$ is an invariant space under the representation U^a.*

b) We have $\mathbf{H}^b = U(g_{ab})\mathbf{H}^a$.

PROOF. a) It is clear that $\mathbf{H}^a$ is a linear space. By the fact that $\lambda^a(\cdot)$ is permissible under G^a we have for $f^a \in \mathbf{H}^a$

$$U^a(g^a)f^a(\phi) = \tilde{f}^a(\lambda^a(\phi g^a)) = \tilde{f}^a((\lambda^a(\phi))g^a) \in \mathbf{H}^a.$$

b) This follows from $U(g_{ab})f^a(\phi) = \tilde{f}^a(\lambda^a(\phi g_{ab})) = \tilde{f}^a(\lambda^b(\phi))$, and conversely.

This is essentially all that is needed to construct the Hilbert space which is so basic to quantum theory. Remember that we have assumed that the data set y^a from experiment a is maximal, which implies that the parameter λ^a is maximal in the sense that it is not a proper function of another parameter that can be connected to a possible experiment. Only one of the different experiments can be performed, which is at the core of the concept of complementarity. Of course, it is possible that experiment a can be divided into several partial experiments which even may be performed at different places, but the parameters of these partial experiments can always be joined in one parameter λ^a.

The point now is that by Theorem 1b), all the Hilbert spaces $\mathbf{H}^a$ are unitarily related. This means that we as a start can pick one fixed, but arbitrary c and define $\mathbf{H} = \mathbf{H}^c$. Then all the other Hilbert spaces are given by

$$\mathbf{H}^a = K^a \mathbf{H} \tag{2}$$

for some unitary operator K^a.

Here we have made a rather concrete construction of the Hilbert space $\mathbf{H}$. In quantum mechanics, the Hilbert space is ordinarily taken as an abstract space. This can be connected to the mathematical fact that all separable Hilbert spaces (with the same number of basis vectors in the finite-dimensional case) are unitarily equivalent.

In Helland (2005b) we proved the following: There is an abstract representation $W(\cdot)$ of the full group G such that $\mathbf{H}$ is an invariant space for $W(\cdot)$. To indicate how $W(\cdot)$ is constructed, take $g_1 \in G^a$ and $g_2 \in G^b$. Then we have

$$W(g_1 g_2) = K^{a\dagger} U^a(g_1) K^a K^{b\dagger} U^b(g_2) K^b. \tag{3}$$

9 States as Hilbert space vectors and density matrices.

Recall that the G^a-invariant space $\mathbf{H}^a$ consists of all functions in $\mathrm{L}^2(\Phi, \nu)$ of the form $\tilde{f}(\lambda^a(\phi))$. From now on we will restrict the theory to the case covered by discrete quantum mechanics, that is, we will assume:

Assumption D. *The parameters λ^a take only a finite number $\{\lambda_k^a\}$ of values.*

Since the range of λ^a constitute an orbit for the group G^a, we take the counting measure as invariant measure. A special basis for the space $\mathbf{H}^a$ is

given by the functions:

$$f_k^a(\phi) = I(\lambda^a(\phi) = \lambda_k^a), \tag{4}$$

where $I(\cdot)$ is the indicator function. These are trivially eigenfunctions for the operators S^a on $\mathbf{H}^a$ defined by

$$S^a \tilde{f}^a(\lambda^a(\phi)) = \lambda^a(\phi)\tilde{f}^a(\lambda^a(\phi)) \tag{5}$$

for functions $f^a(\phi) = \tilde{f}^a(\lambda^a(\phi))$ for which the righthand side of (5) belongs to $\mathbf{H}^a$. Furthermore, the eigenfunctions f_k^a are orthonormal.

We first specialize this to the basic Hilbert space $\mathbf{H} = \mathbf{H}^c$, and use the representation $W(\cdot)$ to construct general state vectors.

It is now convenient to modify this basic Hilbert space. In the present case the group G is finite, hence compact. This implies (Barut and Raczka, Proposition 6, p. 171) that the representation $W(\cdot)$ is unitarily equivalent to a subrepresentation of the regular representation. Thus for some unitary K we make changes $\mathbf{H} \to K\mathbf{H}$, $f_k^c \to v_k^c = Kf_k^c$, $W(g) \to KW(g)K^\dagger$ so that in the new space and for the new operators we still have that $\mathbf{H}$ is invariant for $W(\cdot)$, but we have

$$W(g)v(\phi) = v(\phi g). \tag{6}$$

Also, v_k^c is an eigenvector for $T^c = KS^cK^\dagger$ with eigenvalue λ_k^c.

Definition 2. *With the above notation let $v_k^a = W(g_{ca})v_k^c$ and $T^a = W(g_{ca})T^cW(g_{ac})$.*

Proposition 1. *The vectors v_k^a are eigenvectors of the operators T^a with eigenvalues $\lambda_k^a = \lambda_k^c$.*

Proof. Straightforward.

We will show now is that a considerable subset of all unit vectors of $\mathbf{H}$ are of this form, and that these vectors stand in a natural correspondence with a parameter λ^a together with a fixed value λ_k^a for this parameter. More precisely, these state vectors constitute what is called the generales coherent state vectors in quantum mechanics (Perlomov, 1986 and references there).

Definition 3. *Fix a vector $v_0 \in \mathbf{H}$, and let $W(\cdot)$ be the above representation of the group G. Then every vector of the form $e^{i\alpha}W(g)v_0$, where α is real and $g \in G$, is called a generalized coherent state vector (GCS vectors).*

To arrive at the complete formal structure of quantum mechanics, we make two more assumptions on the representation $W(\cdot)$.

Assumption E. *With v_0 equal to some arbitrary state vector, the set of GCS vectors constitute all the unit vectors in* $\mathbf{H}$.

Without Assumption E we get a quantum mechanics which only is valid for GCS vectors.

With these assumptions, a state vector will correspond to a parameter λ^a and an eigenvalue λ_k^a. Under stronger assumptions, this correspondence will even be one-to-one.

Assumption F. *The mapping $g \to W(g)v_0$ is an injective map from G to the space of unit vectors.*

Theorem 2. *a) Every element of the group G can be written in a unique way as $g = g^c g_{cb}$ for some $g^c \in G^c$.*

b) Fix $v_j^c \in \mathbf{H}$. Then every state vector v_k^a can be written in a unique way as $W(g)v_j^c$ for some $g \in G$. The state v_k^a can be associated with the fact that the parameter $\lambda^a(\phi)$ equals λ_k^a.

c) Under Assumption F, if two state vectors are equal, they correspond to the same parameter λ^a and the same eigenvalue λ_k^a.

The proof of Theorem 2 is deferred to the Appendix.

Physically, the result of Theorem 2 is very important. It means that every vector in $\mathbf{V}$ can be interpreted as a pure state in a very straightforward way: It is equivalent to some question to be determined by an experiment: What is the value of λ^a?, together with some answer: $\lambda^a = \lambda_k^a$.

Equivalently, such a state vector v where the phase factor is ignored, can be represented by a one-dimensional projector: $vv^\dagger$, where $\dagger$ as usual is defined by $v^\dagger u = (v, u)$ for all u.

In practice, the result of an experiment will often be uncertain, given by a probability distribution $\pi(k)$ over $\{\lambda_k^a\}$. In formal quantum mechanics such a mixed state is represented by a density operator

$$\rho = \sum_k \pi(k) v_k^a v_k^{a\dagger}. \tag{7}$$

The set of density operators coincides with the set of positive, self-adjoint operators with unit trace.

In physics, a state may be prepared in several ways, the most straightforward being by doing some experiment and interpreting the result of that experiment as a state; see Section 11 and also Helland (2005b).

10 Transition probabilities.

Assume that we have done an experiment corresponding to the parameter λ^a, and that this has resulted in the crisp value λ^a_k. The information given by this corresponds to the state vector v^a_k. Consider then a new experiment with parameter λ^b. An important result from quantum mechanics, called Born's formula, gives a prior distribution for the last experiment from the information of the first one. This is proved in Helland (2005b) from a symmetry assumption

ASSUMPTION G. *There exists a transition probability* $P(\lambda^b = \lambda^b_i | \lambda^a = \lambda^a_k)$*, and it satisfies*

$$P(\lambda^b = \lambda^b_i | \lambda^a = \lambda^a_k) = P(v^b_i | v^a_k) = P(W(g)v^b_i | W(g)v^b_k).$$

for all $g \in G$.

THEOREM 3. (BORN'S FORMULA) *Under the assumptions above the transition formula is as follows:*

$$P(\lambda^b = \lambda^b_i | \lambda^a = \lambda^a_k) = |v^{a\dagger}_k v^b_i|^2.$$

The proof in Helland (2005b) uses a recent result by Busch (2003), a variant of a classical Theorem by Gleason (1957). This result turns out to have an interesting statistical interpretation.

A straightforward generalization of Born's formula is to the case when the initial information is given by a probability distribution $\pi(k)$ over the parameter values λ^a_k, so that the formal state is given by the density operator (7). Then we have

$$P(\lambda^b = \lambda^b_i | \rho) = v^{b\dagger}_i \rho v^b_i. \tag{8}$$

Born's formula has many physical consequences, some of which will be discussed in the next Section, and others were discussed in Helland (2005b). An interesting point is that it also can be used on ordinary statistical experiments, if there is enough symmetry in the situation.

EXAMPLE 4 (CONTINUED). Recall the 4 experiments with parameters λ^A, λ^B, λ^C and λ^D, each of which can take the values -1 and +1. Assume that we have performed a very accurate experiment with the result that $\lambda^A = +1$. Then it is indicated by a Born formula argument in Helland (2005a) that the prior probability that $\lambda^B = +1$ in a new experiment is given by 1/3, reflecting the fact that μ_A occurs with a minus sign in the formula for λ^B.

11 Inference in statistical language and in Hilbert space language.

The starting point for any inference is a choice of experiment a, the maximal parameter λ^a of that experiment, and the statistical model $p^a(dy^a|\lambda^a)$ for the observations, given the parameter.

In addition to this, for the Bayesian case we need a prior for the parameter, say probabilities $\pi(k)$ for λ_k^a. When a group on the parameter space is defined, which is the case in this paper, we recommend the right invariant measure of the group as a non-informative prior. When λ^a takes a finite number n of values and the group is transitive on the parameter space, this amounts to a probability $1/n$ on each parameter value. In the infinite-dimensional case, we have the usual norming difficulty.

From our point of view, what is new in the quantum theory, is that all this is given a vector or operator representation. First, we let the state vector v_k^a denote that a perfect experiment with maximal parameter λ^a has been performed, and that the result was $\lambda^a = \lambda_k^a$. If our knowledge about λ^a is uncertain, say given by probabilities $\pi(k)$ as above, we represent this knowledge by a density matrix:

$$\rho = \sum_k \pi(k) v_k^a v_k^{a\dagger}. \tag{9}$$

Note that (9) can be given at least two different interpretations. First, $\pi(k)$ can be a prior probability; then this may be called a prior state. The non-informative finite case gives $\rho = n^{-1}I$. Next, $\pi(k)$ can be an aposteriori probability, given some observation; then ρ is an aposteriori state.

The parameter itself can also be represented by an operator:

$$T^a = \sum_k \lambda_k^a v_k^a v_k^{a\dagger} \tag{10}$$

Up to now we have assumed that λ^a is a maximal parameter; this is equivalent to saying that all the eigenvalues are different. However, we have of course also the possibility of defining non-maximal parameters, and corresponding operators defined as in (10) are very often introduced in quantum mechanics. It is a classical result due to von Neumann (1932) that if two operators T^a and T^b commute, then there exits an operator T extending both in the sense that the eigenvectors of T are eigenvectors of both T^a and T^b. In our terminology, if λ^a and λ^b are the corresponding (non-maximal) parameters, then T will correspond to the joint parameter (λ^a, λ^b), which can be associated to a single experiment.

For every operator corresponding to a non-maximal parameter one can find a non-trivial operator commuting to it. On the other hand, if T^a corresponds to a maximal parameter, then only operators corresponding to subparameters commute to it.

PROPOSITION 2. *Let T^a be the operator given by (10). Then there is an operator X commuting with T^a where X corresponds to a parameter which is not a subparameter of λ^a, if and only if λ^a is not maximal.*

PROOF. Assume first that T^a is not maximal; say that the two first eigenvalues λ_1^a and λ_2^a are equal. Then

$$T^a = \lambda u_1^a u_1^{a\dagger} + \lambda u_2^a u_2^{a\dagger} + \sum_{k\geq 3} \lambda_k^a u_k^a u_k^{a\dagger} = \lambda v_1 v_1^\dagger + \lambda v_2 v_2^\dagger + \sum_{k\geq 3} \lambda_k^a u_k^a u_k^{a\dagger}$$

by a rotation to a new orthonormal set. Then every $X = \mu_1 v_1 v_1^\dagger + \mu_2 v_2 v_2^\dagger$ will commute with T^a, and it will not correspond to a subparameter if $\mu_1 \neq \mu_2$.

Assume next that T^a is maximal, that is, that the eigenvalues are different. Assume that there is an $X = \sum_j \mu_j v_j v_j^\dagger$ which commutes with T^a. Expressing the v_j's in terms of the u_k^a's gives $X = \sum_{j,k} \nu_{jk} u_j^a u_k^{a\dagger}$. From this $T^a X = X T^a$ gives $\sum_{j,k} \lambda_j^a \nu_{jk} - \lambda_k^a \nu_{jk} = 0$ $(j \neq k)$. The only solution is $\nu_{jk} = 0$ $(j \neq k)$, so that $X = \sum_k b_k u_k^a u_k^{a\dagger}$. Since the λ_k^a are different, we can always write $b_k = f_k(\lambda_k^a)$. If all b_k are different, this is an operator corresponding to a one-to-one function of λ^a, otherwise to a subparameter.

Born's formula has important consequences for the statistical treatment of the system, since it implies that information can be transferred from one experiment to another. For instance, if we have information in terms of a pure state v_k^a for λ^a and the performs a new experiment b, then this gives a prior for the new experiment with expectation

$$\mathrm{E}(\lambda^b | \lambda^a = \lambda_k^a) = v_k^{a\dagger} T^b v_k^a.$$

For other consequences of Born's formula and standard inference theory, we refer to Helland (2005b).

It is of particular interest that the process of observation changes our information abruptly, and hence changes the state. Quantitatively, this can be expressed by using Bayes formula. In the quantummechanical literature the corresponding result is called von Neumann's projection postulate, and has caused much discussion. In von Neumann (1932) and elsewhere one has tried to study the phenomenon deeper by introducing a new model

which incudes the measurement apparatus. This is of importance for the consistency of the theory, but it has not much interest from an applied point of view.

12 Model reduction in statistical practice.

Let us go back to the issue raised in the introduction: Every model has its limitation in the sense that while it gives a language under which to analyze the data under given circumstances, a detailed ‘true’ model can never be found. Then it should be clear that the degree of sophistication of the model chosen may depend upon the amount of data that is available for the statistical analysis. Sometimes it might happen that it makes sense to use complementary models and complementary parameters to answer different, complementary questions. An example of the latter is when different sets of orthogonal contrasts are used in analysis of variance. However, in this Section we will concentrate on the situation where one single model with a single set of parameters λ is to be chosen.

In the statistical literature there is much discussion about which criteria that should be used to select a model, but how to select the set of potential models to choose from, is not much debated. The examples below will show that in this process our criterion A can be very useful: The parameter set of the model should constitute an orbit or a set of orbits of a group which for some reasons may be connected to the model.

Example 5. Consider a single series of measurements $y_1, y_2, \ldots, y_n$. Assume to start with a very rich model with a parameter set ϕ which is invariant under the group G of location and scale change. If this model is rich enough and if n only has a reasonable size, it may be very difficult to do parametric inference from this.

Any reduced model should be an orbit or a set of orbits under G. An obvious candidate is then the class of normal distributions. This may give a partial explanation (in addition to completely different arguments given in the literature) why the assumption of normality is so useful in applied statistics.

Example 6. Consider a two-sample t-test situation: $y_{11}, \ldots, y_{1n_1}$ are independent $N(\mu_1, \sigma_1^2)$, while $y_{21}, \ldots, y_{2n_2}$ are independent $N(\mu_2, \sigma_2^2)$. Let G consist of the transformations $y_{1j} \rightarrow a_1 + by_{1j}$, $y_{2j} \rightarrow a_2 + by_{1j}$. These transformations make the model assumptions invariant. It is important to

use the same b on both samples, for we want the parameter $\mu_1 - \mu_2$ to be permissible.

Again, any reduced model should be an orbit or a set of orbits under G. It is easy to check that the orbits are given by $\sigma_1/\sigma_2 =$ constant. A particular and very much used model simplification is given by the orbit $\sigma_1 = \sigma_2$. As in the previous example, the truth of such a model reduction assumption can never be proved; one must only check that it is not inconsistent with data.

EXAMPLE 7. Look at a multiple regression model with dependent variable y and explanatory variables $x_1, x_2, \ldots, x_p$. Assume that the explanatory variables the different units. Then a relevant group is the group G of separate scale transformations: $x_i \rightarrow b_i x_i; i = 1, \ldots, p$, where one must have $b_i \neq 0$. This induces a transformation of the regression parameters given by $\beta_i \rightarrow \beta_i/b_i$. The range of each single β_i then has two orbits: $\{\beta_i : \beta_i \neq 0\}$ and $\beta_i = 0$. Every subset of the regression parameter space obtained by putting some β_i's equal to 0 will then be a set of orbits under the group G. These are just the subsets of the parameter space that are ordinarily used in model selection in regression models.

EXAMPLE 8. Again, consider the multiple regression model, but let now all the explanatory variables have the same units. Then a much larger transformation group can be considered, for instance the affin group: $\beta \rightarrow A\beta$, where $\beta = (\beta_1, \ldots, \beta_p)'$ and A is any non-singular matrix. This group is not very interesting for model reduction purposes, however, since it is transitive on the parameter space.

A more interesting group G is obtained if we limit A to always be an orthogonal matrix. This group is interesting since common biased regression methods like principal component regression and ridge regression are *equivariant* under orthogonal transformations: The estimated regression vectors transform in the same way as the parameters. Under transitivity and quadratic loss, the best equivariant estimator is equal to the Bayes estimator with right invariant prior.

In Helland (2001) it was shown that the partial least squares regression model of Helland (1990) constitute a set of orbits under the orthogonal transformation group. Maximum likelihood under this reduced model was discussed in Helland (1992), but is not practical, since the model still has too many parameters. It is conjectured that the Bayes estimator under right invariant prior supplemented with maximum likelihood or restricted maximum likelihood for the orbit index may be developed into a practical method

which may improve the partial least squares method used by chemometricians. If this can be done, it is also conjectured that the solution will provide a good regression method for the case with many explanatory variables.

Much research remains to be done on model reduction in the multiple regression model and in related models. This research may be important, since overparametrized models are now being proposed in very many different areas. If the link towards quantum mechanics is being accepted, there is a chance that one some day may benefit from parts of the research that has been done on numerical methods connected to quantum theory.

Example 9. Look at a simplified description of a design of experiment process. We will let the reader judge the closenes of the theory to what has been discussed earlier in this paper.

Consider a set Z of potential experimental units for some experiment; this set can be finite or infinite, and one may even consider an uncountable number of units. For each given $z \in Z$, let y_z be some potential response variable, and let μ_z be the expectation of y_z for the case where no treatment is introduced. One may also have a set T of potential treatments which can be applied to each unit. Let μ_{tz} be the expectation of y_z, given z, when treatment t is applied to z, and define $\lambda_{tz} = \mu_{tz} - \mu_z$. Assume for simplicity that the y_z's are independent with a constant variance σ^2. Let η_z denote other parameters connected to the unit z.

In this situation it is natural to call $\phi = (\{\mu_z, \eta_z; z \in Z\}, \{\lambda_{tz}; t \in T, z \in Z\}, \sigma^2)$ a total parameter for the system and $\Phi = \{\phi\}$ the total parameter space. Note that ϕ of course is not estimable in any conceivable experiment; nevertheless it is a useful conceptual quantity.

Let G be a transformation group defined on Z. This will induce a group on Φ.

Now for the experiment itself select a finite subset Z_0 of Z. We will assume for simplicity that G is so large that the full permutation group G_0 on Z_0 is a subgroup of G.

We will also assume that Z_0 is selected from Z by some random mechanism with the property that $\lambda_t = \mathrm{E}(\lambda_{tz}|t)$, expectation over this selection mechanism, is independent of the selected z. Then we will have for a given selected unit $z \in Z_0$ that

$$\mathrm{E}(y_z|t) = \mu_z + \lambda_t.$$

This is one way to express the well known unit/treatment additivity, which is considered by Bailey (1981, 1991) and others to be crucial for having a consistent approach to the design of experiments.

From this point on Bailey (1981) introduces an eight-stage experimental design theory, and this theory is developed further in Bailey (1991). We will only mention very briefly a few main points of this theory, referring to these and related papers for details.

Block structure is an important aspect of experimental design theory and practice: Similar units are taken together in one block to enhance efficiency. This topic has many important facets, like Latin squares, split plot blocking, incomplete blocks and so on. From a group theoretical point of view, the main point is that the block structure determines the group used for randomization: For a selected experiment a, use for randomisazion the largest subgroup G^a of G_0 which respects the block structure of that experiment: If the units z_1 and z_2 are in the same block, then $z_1 g$ and $z_2 g$ should be in the same block for all $g \in G^a$. The unit (names) are then randomized according to this group. This randomization also has connections to the allocation of treatments.

Assuming that G^a is transitive, Bailey (1991) proves the following: After randomization, y_z (overusing this symbol slightly) has an expectation which only depends upon the treatment $t(z)$ given to z, and a covariance matrix C satisfying

$$C(z_1, z_2) = C(z_1 g, z_2 g), \tag{11}$$

for $z_1, z_2 \in Z_0$ and $g \in G^a$. Using this, Bailey (1991) introduces the *strata*, which are the eigenspaces of C, and which also are invariant spaces under (a representation of) the group G^a. The important practical point is that these give the lines of the (null) analysis of variance for the experiment, both in simple and in complicated cases.

13 Discussion and conclusions.

Several conclusions seem to be possible to draw from the discussion above both for the field of quantum mechanics and the field of statistics. Both disciplines have had great empirical success, so it is very natural to understand that many scientists are sceptical towards radical changes in the foundation that they are used to have as a point of departure. On the other hand, if one believes the message of this paper, there seems to be a logical connection between the two disciplines, and this should in principle imply that it should be possible to develop at least some common attitudes to the process of data analysis. The distance between the two communities is rather large today. There is little reasons to claim - in fact it is probably untrue - that all of this

difference has a cultural origin. But it seems highly likely that some cultural differences exist, and that this may be an opportunity for scientists to learn from each other. The discussion given below on this and related issues must be taken as tentative. One may hope that the relationship between the different sciences will be better understood as time goes on.

Some remarks on quantum physics:

1) For me it seems clear that the group representation approach together with the question-and-answer interpretation has the potential of becoming a more natural way to introduce the theory for discrete quantum mechanics than the ordinary formal approach. Still, the formal apparatus may be the best one when doing calculations, at least in many instances.

2) The concept of underlying model and the total parameter ϕ has been exemplified for the spin case, but not in other cases in the treatment above. More research on this remains to be done. It may be that one in certain cases, say when studying the particle aspect versus the wave aspect of an electron, will find it convenient to use a few, complementary models instead of one. The goal should be one underlying model, however.

3) The electron spin example may be extended to include general spin as follows: Let $\lambda^a = (|\phi|, \tau^a)$, where τ^a now is the spin component in direction a. Let the group G consist of rotations of ϕ together with norm changes.

4) A natural way to model a free particle, a particle in a box or a particle in the double slit experiment is to let $\phi^a = (\xi^a, \pi^a)$, where ξ^a is the position of the particle and π^a is its momentum. A natural proposal of a group may be to consider the translations together with the Galilei transformations.

5) The treatment of quantum mechanis in this paper is far complete. Important missing parts are the arguments behind Planck's constant and the Schrödinger equation. We hope to treat both themes in Helland (2005c). Other important parts are discussion of the interaction with the measuring apparatus, interference, continuous statistical parameters and relation to quantum mechanics based upon C*-algebra, relativistic quantum mechanics, field theory and so on.

6) In particular, we have not included any discussion of entanglement; in our approach a tempting model of this is that of two subsystems connected to the same total parameter. Consider for instance two electrons in a joint state determined by a total spin vector ϕ for the first electron and $-\phi$ for the second electron. If a measurement in the direction a is done on the first electron with the result $+1$, then a measurement in direction a on the second electron will be -1 always. In general the correlations between spin measurements on the two electrons are stronger than what can be explained

by a simple local hidden-variable model. One way to show this is through the well-known Bell inequality, which has to do with the situation where one has the choice between measurements in the the directions a and b for one particle and c and d for the other particle.

Bell's inequality is derived from an algebraic relation found under the assumption of a local hidden variable model, and then the expectation over the variables of this relation is taken. Here we only remark that the situation given really involves 4 different experiments, and that according to ordinary statistical properties, expectations should be taken conditionally, given the experiment chosen. If this is done, the proof of Bell's inequality breaks down.

For some relevant considerations related to entanglement, see Helland (2005d).

7) The approach here also has a saying on the so-called paradoxes of quantum mechanics. Consider for example the famous Schrödinger's cat: A cat is enclosed in a box together with a poison capsule and a radioactive particle. When the particle disintegrates, the capsule breaks, and the cat dies. In ordinary quantum mechanics it has been a puzzle that states for this system can be created where the cat is partly alive, partly dead. From our point of view this seems to be no big problem: The relevant states are simply connected to a maximal set of questions which is not concerned with the death status of the cat.

Some remarks on statistics:

1) The result of this paper may be taken as an argument against an attitude where statistics is taught only as a mathematical deduction from probability theory and probability models. As I see it, other aspects of applied inference are very important, including choice of experimental question, choice of model, and symmetry. Our teaching - and also our theoretical research - should take this as a point of departure also. More emphasis should probably be placed upon learning from examples and from real problems in our teaching, but I do not say that this is easy to follow up.

2) Much emphasis should be given to simple models that can be seen as orbits or sets of orbits of some underlying natural group. This includes for instance simple analysis of variance models. In introductory courses, at least for users, these models could be introduced without using any group theory.

3) The parameter of a statistical model should not only be taken as an index describing a class of probability measures. Parameters most often have important direct interpretations.

4) Parameters make sense also in cases where there is no real or conceptual population with respect to which asymptotic analysis can be made.

5) Model development is important; an initial model may often be too rich to analyse statistically.

6) Model reduction is also important, but not only the aspect of determining a criterion for the reduction process. Equally important might be to find a suitable set of candidate models to which to reduce.

7) A reduced model should not necessarily be looked upon as an approximation.

8) Bayesian inference is useful. A reasonable non-informative prior is found from the righthand prior of the underlying group. In this case a close connection to non-Bayesian inference can be found.

9) More emphasis should be placed on the interplay between experimental design and inference; between question and answer. Sometimes it may be useful to ask several complementary questions, even use complementary models in the same situation. For observational data a similar emphasis can be argued for, taking into account contrafactual questions concerning the mechanism behind the data generation.

10) There is much research that remains to be done on overparametrized models. Some information on where to search for model reduction may be found from the methods proposed by applied researchers. The partial least squares methods of chemometricians (example 8) is an illustration of this issue.

Appendix.

PROOF OF THEOREM 2. a) Every group element $g \in G$ is of the form $g^c g_{cb}$ for some g^c and some b. For instance, let $g = g_1^a g_2^b g_3^d$for $g_1^a \in G^a$, $g_2^b \in G^b$ and $g_3^d \in G^d$. Then

$$g = g_1^a g_2^b (g_4^b g_{bd}) = g_1^a g_5^b g_{bd} = g_1^a (g_6^a g_{ab}) g_{bd} = g_7^a g_{ad} = g_8^c g_{ca} g_{ad} = g_8^c g_{cd}.$$

The decomposition $g = g^c g_{cb}$ is unique since the transformations g_{cb} between parameters are unique and g^c only serves to permute the eigenvalues.

b) We have $W(g^c g_{cb}) v_j^c(\phi) = v_j^c(\phi g^c g_{cb})$. Furthermore, v_j^c is associated with the fact that the parameter $\lambda^c(\phi)$ equals λ_j^c. Also, the function of g^c is the change j to another value k, and $\lambda^c(\phi g_{cb}) = \lambda^b(\phi)$. Transforming unitarily to the space $\mathbf{H}^a$ and back again, this means that the vector $v_k^a = W(g^c g_{cb}) v_j^c$ has the interpretation that the parameter $\lambda^a(\phi)$ equals λ_k^a.

c) Assume that $W(g_1)v_j^c = W(g_2)v_j^c$. Then by Assumption F we have that $g_1 = g_2$, which means that they have the same representation of the form $g^c g_{cb}$. But then the statement $\lambda^c = \lambda_j^c$ is transformed to the same statement $\lambda^a = \lambda_k^a$.

Acknowledgements.

I am grateful to Trond Reitan for detailed comments to an earlier version of this paper. Comments on the paper by Fredrik Dahl and Anders Rygh Swensen are also acknowledged.

References.

Bailey, R.A. (1981) A unified approach to design of experiments. *J. R. Statist. Soc.* A, 144, 214-223

Bailey, R.A. (1991) Strata for randomized experiments. *J. R. Statist. Soc.* B, 53, 27-78.

Barut, A.O. and Raczka, R. (1977) *Theory of Group Representations and Applications.* Warszawa: Polish Scientific Publishers.

Barndorff-Nielsen, O.E., Gill, R.D. and Jupp, P.E. (2003) On quantum statistical inference. *J. R. Statist. Soc.* B, 65, 775-816.

Bell, J.S. (1966) On the problem of hidden variables in quantum theory. *Rev. Mod. Phys.*, 38, 447-452.

Bohm, D. (1952) A suggested interpretation of the quantum theory in terms of "hidden" variables, I. *Phys. Rev.*, 85, 166-179.

Bovens, L. and Hartmann, S. (2003) *Bayesian Epistemology.* Oxford: Claredon Press.

Bozdogan, H. (1988) Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. *Psychometrika*, 52, 345-370.

Breiman, L. (2001) Statistical modelling: The two cultures. *Statistical Science*, 16, 199-231.

Burnham, K.P. and Anderson, D.R. (1998) *Model Selection and Inference. A Practical Information-Theoretical Approach.* New York: Springer-Verlag.

Busch, P. (2003) Quantum states and generalized observables: A simple proof of Gleason's theorem. *Phys. Rev. Letters*, 91, 12.

Diaconis, P. (1988) *Group Representations in Probability and Statistics.* IMS Lecture Notes. Monograph Series, Vol. 11. Hayward, California: Institute of Mathematical Statistics.

Drum, M. and McCullagh, P. (1993) Comments on "Regression models for discrete longitudinal responses". *Statistical Science*, 8, 300-301.

Fuchs, C.A. (2002) Quantum mechanics as quantum information (and a little more). http://arxiv.org/abs/quant-ph/0205039.

Gleason, A. (1957) Measures on closed subspaces of a Hilbert space. *J. Math. Mech.*, 6, 885-893.

Helland, I.S. (1990) Partial least squares regression and statistical models. *Scand. J. Statist.*, 17, 97-114.

Helland, I.S. (1992) Maximum likelihood regression on relevant components. *J. R. Statist. Soc.* B, 54, 637-647.

Helland, I.S. (2001) Reduction of regression models under symmetry. Invited contribution to: Viana, M. and D. Richards (ed.) *Algebraic Methods in Statistics.* Contemporary Mathematics Series. Providence, Rhode Island: American Mathematical Society.

Helland, I.S. (2004) Statistical inference under a fixed symmetry group. *Int. Statist. Rev.*, 72, 409-422.

Helland, I.S. (2005a) Quantum theory as a statistical theory under symmetry. In Foundations of Probability and Physics - 3. (ed. A. Khrennikov) AIP Conference Proceedings. Vol. 750. Melville, New York: American Institute of Physics. Revised version on http://arxiv.org/abs/quant-ph/0411174.

Helland, I.S. (2005b) Extended statistical modelling under symmetry; the link towards quantum mechanics. To appear in *Ann. Statist.* Also available on http://arxiv.org/abs/quant-ph/0503214.

Helland, I.S. (2005c) Quantum theory as a statistical theory under symmetry, complementarity, and time evolution. Manuscript.

Helland, I.S. (2005d) Quantum phenomena may have macroscopic counterparts. Submitted.

James, G. and Liebeck, M. (1993) *Representation and Characters of Groups.* Cambridge: Cambridge University Press.

Perelomov, A. (1986) *Generalized Coherent States and Their Applications.* Berlin: Springer-Verlag.

Schweder, T. and Hjort, N.L. (2003) Frequentist analogues of priors and posteriors. In *Econometrics and the Philosophy of Economics. Theory-Data Confrontations in Economics.* (ed. B.P. Stigum). Princeton: Princeton University Press.

Taraldsen, G. (1995) Om kvantemekanikk og kunsten å gjette. (In Norwegian.) *Symmetri* 2, 23-29.

von Neumann. J. (1932) *Mathematische Grundlagen der Quantenmechanik.* Berlin: Springer.